

# **PREDICCIÓN DE PRECIOS Y DEMANDAS DE ENERGÍA ELÉCTRICA BASADA EN MODELOS DE REGRESIÓN HARMÓNICA DINÁMICA**

**Juan R. Trapero, Diego J. Pedregal, Fausto P. García Márquez**

Área de Organización de empresas. Escuela Técnica Superior de Ingenieros Industriales de Ciudad Real. Campus de Ciudad Real, 13071 Ciudad Real. [JuanRamon.Trapero@uclm.es](mailto:JuanRamon.Trapero@uclm.es), [Diego.Pedregal@uclm.es](mailto:Diego.Pedregal@uclm.es), [FaustoPedro.Garcia@uclm.es](mailto:FaustoPedro.Garcia@uclm.es).

## **Resumen**

*Los cambios experimentados por los mercados eléctricos recientemente en numerosos países ha provocado la aparición de nuevos retos sobre la predicción de las series temporales de precios y demandas de energía eléctrica. En el presente trabajo se propone una herramienta basada en un marco de espacio de los estados utilizando un modelo de componentes no observables. Esta técnica presenta ventajas en dos frentes, el técnico y el de gestión. Las ventajas técnicas producen predicciones muy competitivas basadas en el carácter adaptativo y recursivo del modelo y el algoritmo de estimación utilizado. Por otro lado, las ventajas de gestión vienen de la mano del marco del espacio de los estados empleado, el cual no sólo ofrece la predicción sino que también proporciona la posibilidad de extraer de forma explícita la tendencia de los precios del mercado, su comportamiento estacional y la componente irregular.*

**Palabras clave:** Regresión harmónica dinámica, espacio de los estados, mercados eléctricos.

## **1. Introducción**

La liberalización del mercado eléctrico que se está llevando a cabo recientemente en numerosos países ha propiciado el aumento de interés sobre las herramientas de predicción. Sobre el estudio de este problema podemos encontrar una gran cantidad de literatura. Entre las soluciones más conocidas nos encontramos desde los clásicos modelos ARIMA por ejemplo Contreras *et al.* (2003) hasta el uso de redes neuronales, ver por ejemplo Kumluca y Erkmen (2003). Los modelos de componentes no observables (*Unobserved Components, UC*) se encuentran bien situados para su aplicación en las series estudiadas en este artículo, ver Trapero (2003). Otros métodos muy citados son: método Bayesiano en West y Harrison (1989); espacio de los estados en el dominio del tiempo en Harvey (1989), Pedregal y Young (2002); espacio de los estados en el dominio de la frecuencia en Young *et al.* (1999); también métodos basados en reducciones del modelo, Kumluca y Erkmen (2003). Todos estos métodos, aunque parten de una descomposición en componentes teóricos semejantes, tienen puntos de vista completamente diferentes en referencia a su forma de realizar la identificación y estimación.

A continuación vamos a hacer un resumen de cómo se ha organizado el artículo. El segundo apartado se proporciona una introducción al espacio de los estados. En este apartado se presentarán los modelos de componentes no observables y como particularización el modelo DHR. La aplicación a las series temporales de demandas y precios de energía de diferentes mercados se desarrollará en el tercer apartado. En el cuarto apartado se muestra la

implementación de nuestro modelo en un paquete informático denominado forecastool que nos permite utilizar el modelo estudiado de una forma gráfica e intuitiva. Por último se muestran las principales conclusiones en el quinto apartado.

## 2. Modelo de espacio de los estados

Los modelos de espacio de los estados se componen por dos conjuntos de ecuaciones:

$$\begin{aligned} \text{Ecuaciones de estado:} & \quad \mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{w}_t \\ \text{Ecuaciones de observación:} & \quad z_t = \mathbf{H}_t\mathbf{x}_t + v_t \end{aligned} \quad (1)$$

donde  $z_t$ ,  $t=1,2,\dots,N$  es la serie temporal escalar;  $\mathbf{x}_t$  es un vector estocástico de dimensión  $nx1$ ;  $\mathbf{w}_t$  y  $v_t$  son vectores de ruido de dimensión  $nx1$  y  $1x1$  respectivamente; y  $\mathbf{A}$  y  $\mathbf{H}_t$  son las matrices del sistema y de salida respectivamente. Es importante destacar el subíndice de la matriz de salida  $\mathbf{H}_t$  el cual indica su dependencia temporal. Como hipótesis de partida decimos que  $\mathbf{w}_t$  es un ruido Gaussiano con vector de medias cero y matriz de covarianzas  $\mathbf{Q}^*$ . De forma similar, el ruido escalar  $v_t$  es Gaussiano con media cero y varianza  $\sigma_v^2$ . Además ambos ruidos son independientes. Las matrices  $\mathbf{A}$ ,  $\mathbf{H}_t$ ,  $\mathbf{Q}^*$  y  $\sigma_v^2$  son conocidas o han sido estimadas previamente

La formulación utilizando el modelo de espacio de los estados nos permite manejar sistemas dinámicos lineales no estacionarios. Si las perturbaciones en el modelo son gaussianos el *filtro de Kalman* (KF) y el algoritmo de *Suavizado de intervalo fijo* (FIS) producen las estimaciones óptimas en el sentido de minimizar los errores cuadráticos medios. La versión de ambos algoritmos usados en este artículo aparecen en el apéndice. Debido a que las predicciones se han de presentar en el mercado diario (en el caso del operador de mercado español) cada 24 horas no es necesario que trabajemos en línea, por lo que aprovechamos esto para utilizar un algoritmo de suavizado como es el FIS. La aplicación de los algoritmos recursivos KF/FIS necesitan el conocimiento de todas las matrices del sistema  $\mathbf{A}$ ,  $\mathbf{H}_t$ ,  $\mathbf{Q}^*$ , y  $\sigma_v^2$ .

### 2.1. Modelos de componentes no observables y regresión armónica dinámica.

Una regresión armónica dinámica es una particularización del conjunto de modelos de componentes no observables que adopta la siguiente forma:  $v_t$

$$z_t = \sum_{i=0}^R [a_i \cos(\omega_i t) + b_i \sin(\omega_i t)] + v_t \quad (2)$$

donde  $z_t$  representa la serie temporal escalar y  $v_t$  es ruido blanco de la misma forma que en (1);  $\omega_i$ ,  $i=1,2,\dots,R$  hace referencia al conjunto de la frecuencia fundamental y sus armónicos que contiene la serie temporal; por último  $a_i$  y  $b_i$  son los parámetros variantes en el tiempo (TVP) estocásticos, cuya naturaleza estocástica sigue un paseo aleatorio generalizado, tal que:

$$\begin{pmatrix} x_{it+1} \\ x_{it+1}^* \end{pmatrix} = \begin{pmatrix} \alpha & \beta \\ 0 & \gamma \end{pmatrix} \begin{pmatrix} x_{it} \\ x_{it}^* \end{pmatrix} + \begin{pmatrix} w_{it} \\ w_{it}^* \end{pmatrix} \quad (3)$$

donde  $x_{it}$  representa cualquiera de los TVP's en (2);  $x_{it}^*$  es un estado adicional en el modelo sin una clara interpretación, en general;  $\alpha$ ,  $\beta$  y  $\gamma$  son parámetros adicionales;  $w_{it}$  y  $w_{it}^*$  son iid ruidos Gaussianos. Este modelo general abarca casos particulares tanto como otros generalmente usados en la literatura como el IRW (*Integrated Random Walk*,  $\alpha = \beta = \gamma = 1$ ;  $w_{it} = 0$ ); el RW (*Random Walk*:  $\alpha = 1$ ;  $\beta = \gamma = 0$ ;  $w_{it}^* = 0$ ), el LLT (*Local Linear Trend*:  $\alpha = \beta = \gamma = 1$ ) y el DT (*Damped Trend*:  $\alpha = \beta = 1$ ).

La selección de las frecuencias relevantes en una serie temporal es un tema de identificación. Afortunadamente para las series temporales estudiadas en este artículo correspondiente a las demandas y precios de energía eléctrica poseemos un conocimiento a priori de esas frecuencias fundamentales, por el conocimiento del problema en concreto. En nuestro caso particular los periodos estacionales importantes corresponden tanto a  $T=24$  observaciones, que representa una estacionalidad diaria, y  $T=24*7=168$  observaciones, que representa una estacionalidad semanal. Sin embargo las series temporales que manejamos no son ondas senoidales puras, por tanto contienen armónicos y es en la selección de qué armónicos introducir al modelo donde mayores problemas nos encontramos. Dado que el proceso no es estacionario, el AR-espectro obtenido es mezcla de tendencia, componente estacional y ruido. Esto dificulta la identificación de los armónicos que aportan mayor información del sistema mediante algún tipo de test estadístico. Para solventar este problema y en el caso particular de las series temporales obtenidas en el mercado eléctrico, podemos proceder de la siguiente manera: sea el parámetro de diseño  $\gamma$  tal que:

$$\phi(\omega_k) > \gamma \quad (4)$$

donde  $\phi(\omega_k)$  es una medida del contenido en frecuencia de la serie temporal para una determinada frecuencia  $\omega_k$ . Los armónicos de las frecuencias fundamentales que incluiremos en (2) serán aquellos que cumplan (4). En nuestro caso concreto hemos tomado  $\phi(\omega_k)$  como la amplitud obtenida por el AR-espectro en la frecuencia  $\omega_k$ . El parámetro  $\gamma$  lo hemos obtenido empíricamente para el caso particular de estas series, llegando a un compromiso entre la información aportada por los armónicos y la sensibilidad al ruido, obteniendo un valor de  $\gamma = 0.4$

La no estacionariedad de la serie puede ser incorporada en el modelo mediante los TVP's  $a_{it}$  y  $b_{it}$  ya que estos no son estacionarios en sí mismos. El modelo (2) puede ser escrito explícitamente en términos de componentes no observables, tal que:

$$z_t = T_t + S_t + v_t \quad (5)$$

donde  $T_t$ ,  $S_t$  y  $v_t$  pueden ser interpretados como un término de tendencia estocástica, un componente estacional estocástico y un componente irregular, respectivamente.

El modelo DHR en (2) posee una serie de TVP's ( $a_{it}$  y  $b_{it}$ ) los cuales siguen una evolución del tipo mostrado en (3). Las ecuaciones de estado del DHR estarán formadas por un IRW para modelizar la tendencia y un RW para modelizar las amplitudes de los armónicos. Las

frecuencias que tendremos en cuenta en (2) corresponderán a los periodos seleccionados de acuerdo con (4) . Por tanto la ecuación de observación es:

$$z_t = \sum_{i=0}^N [a_{it} \cos(\omega_i t) + b_{it} \sin(\omega_i t)] + v_t \quad (6)$$

donde N es el número de frecuencias seleccionadas que incluimos en (2). La evolución de los TVP's vendrán dados por las siguientes ecuaciones de estado:

$$\begin{bmatrix} a_{0t+1} \\ a_{0t+1}^* \\ a_{1t+1} \\ b_{1t+1} \\ \vdots \\ a_{it+1} \\ b_{it+1} \\ \vdots \\ a_{Nt+1} \\ b_{Nt+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & & \cdots & & 0 & 0 \\ 0 & 1 & 0 & & \cdots & & 0 & 0 \\ 0 & 0 & 1 & 0 & & \cdots & & \\ & & & 0 & 1 & 0 & & \\ & & & & 0 & \ddots & \ddots & \\ \vdots & \vdots & & & & \ddots & 1 & 0 \\ & & & & & & 0 & 1 & \ddots \\ & & & & & & & \ddots & \ddots & 0 & 0 \\ 0 & 0 & & & \cdots & & 1 & 0 \\ 0 & 0 & & & \cdots & & 0 & 1 \end{bmatrix} \begin{bmatrix} a_{0t} \\ a_{0t}^* \\ a_{1t} \\ b_{1t} \\ \vdots \\ a_{it} \\ b_{it} \\ \vdots \\ a_{Nt} \\ b_{Nt} \end{bmatrix} + \begin{bmatrix} 0 \\ \omega_{0t}^* \\ \omega_{a1t} \\ \omega_{b1t} \\ \vdots \\ \omega_{ait} \\ \omega_{bit} \\ \vdots \\ \omega_{aNt} \\ \omega_{bNt} \end{bmatrix} \quad (7)$$

$$z_t = \left[ 1 \quad 0 \quad \cos\left(\frac{2\pi}{T_1}t\right) \quad \sin\left(\frac{2\pi}{T_1}t\right) \quad \cdots \quad \cos\left(\frac{2\pi}{T_i}t\right) \quad \sin\left(\frac{2\pi}{T_i}t\right) \quad \cdots \quad \cos\left(\frac{2\pi}{T_N}t\right) \quad \sin\left(\frac{2\pi}{T_N}t\right) \right] x_t + v_t$$

$$Q^* = \text{diag}\left(0 \quad \sigma_0^2 \quad \sigma_1^2 \quad \sigma_1^2 \quad \cdots \quad \sigma_i^2 \quad \sigma_i^2 \quad \cdots \quad \sigma_N^2 \quad \sigma_N^2\right)$$

donde  $\text{diag}(\bullet)$  es una matriz diagonal cuyo elementos pertenecen a la diagonal principal. El modelo (7) puede ser reparametrizado convenientemente sin perder generalidad mediante una transformación de la matriz de covarianza  $Q^*$  en un matriz de ratios conocida como la matriz NVR (*Noise Variance Ratios*), tal que:

$$Q = \text{diag}\left(0 \quad \frac{\sigma_0^2}{\sigma_v^2} \quad \frac{\sigma_1^2}{\sigma_v^2} \quad \frac{\sigma_1^2}{\sigma_v^2} \quad \cdots \quad \frac{\sigma_i^2}{\sigma_v^2} \quad \frac{\sigma_i^2}{\sigma_v^2} \quad \cdots \quad \frac{\sigma_N^2}{\sigma_v^2} \quad \frac{\sigma_N^2}{\sigma_v^2}\right) = \\ = \text{diag}\left(0 \quad NVR_0 \quad NVR_1 \quad NVR_1 \quad \cdots \quad NVR_i \quad NVR_i \quad \cdots \quad NVR_N \quad NVR_N\right) \quad (8)$$

El parámetro  $\sigma_v^2$  es la varianza del ruido observado, y lo utilizamos en el modelo DHR (2) para normalizar las varianzas, sin que influya sobre la esencia del modelo. Otra restricción habitual en el modelo es que las varianzas o NVR's correspondientes a los términos seno y coseno de cada frecuencia sean los mismos. Comparando (7) con el modelo general (1) podemos identificar claramente las matrices y estados del sistema, así como el vector de ruidos y las matrices de covarianzas. Una vez identificados todos los elementos del sistema, los algoritmos KF/FIS producirían estimaciones óptimas de sus estados. El problema que nos encontramos reside en que es necesario estimar a priori una serie de parámetros desconocidos, que son los NVR en la matriz  $Q$ . Normalmente el método más ampliamente utilizado para estimar los NVR es por máxima verosimilitud (ML) en el dominio del tiempo a causa de sus buenas propiedades teóricas en una gran cantidad de aplicaciones. Sin embargo, este método no es el más apropiado cuando el número de parámetros a estimar es alto. En nuestro caso ocurre este fenómeno por eso debemos acudir a otras alternativas. Afortunadamente, la

estimación en el dominio de la frecuencia proporciona una alternativa adecuada a este problema. Por tanto, podemos resolverlo ajustando las curvas procedentes del (pseudo) espectro empírico de la serie temporal y del (pseudo) espectro del modelo. Una versión completa de este algoritmo de estimación se puede ver en Young *et al.* (1999). A continuación vamos a establecer los pasos a seguir para realizar la estimación en el dominio de la frecuencia:

1. Estimar el espectro empírico de los datos.
2. Establecer el modelo DHR con los picos identificados en (4).
3. Calcular mediante ML en el dominio de la frecuencia los NVR.
4. Usar las estimaciones en 3 para ejecutar los algoritmos KF/FIS, para realizar la predicción, interpolación, estimación de la componente de tendencia o de la componente estacional, etc.

### 3. Caso de estudio

El modelo de regresión armónica dinámica lo hemos aplicado a los precios y demandas del mercado eléctrico español, cuyos datos se han obtenido de la página web del operador de mercado ([www.omel.es](http://www.omel.es)). También se ha utilizado para la predicción de precios del mercado PJM interconnection de los EEUU, para corroborar la robustez de nuestro modelo. Todas estas operaciones se realizarán mediante la toolbox de Matlab llamada Captain, ver Young *et al.* (2003).

El ejercicio predictivo consiste en realizar la predicción de 24 horas hacia delante durante la última semana de cada mes del año, teniendo en cuenta la información correspondiente de todo el mes. Los resultados obtenidos con el modelo DHR se compararán con los de un modelo ingenuo (NAIVE), consistente en tomar como predicción los valores correspondientes a la semana anterior directamente. Las comparaciones de las predicciones de ambos modelos se llevarán a cabo de acuerdo a las siguientes medidas estadísticas de error: Error absoluto medio normalizado (NMAE) y su desviación típica (STD) tal que:

$$NMAE = \frac{1}{24} \frac{\sum_{l=1}^{24} |\hat{z}_{t+l} - z_{t+l}|}{\bar{z}_{t+l}} \cdot 100 \quad (9)$$

donde  $\hat{z}_{t+l}$  y  $z_{t+l}$  son los valores correspondientes a la predicción de la demanda y su valor real respectivamente, adelantados  $l$  pasos hacia delante respecto al tiempo  $t$ . El valor  $\bar{z}_{t+l}$  será la media muestral de los  $l$  valores reales utilizados en la predicción. Las razones de normalizar el error absoluto con la media muestral son: 1) Poder comparar los errores cometidos tanto en la serie de demandas como en la de los precios y 2) evitar valores de errores sobrevalorados para valores de precios cercanos a cero. En la Tabla 1 mostramos los errores cometidos por nuestro modelo DHR al realizar la predicción de la última semana de cada mes en el año 2002 para la serie temporal formada por las demandas de energía del mercado español. Podemos observar que la medida de error es el promedio de (9) a lo largo de una semana (es decir, hemos realizado la media de los 7 NMAE obtenidos en la predicción de cada día de dicha semana y lo hemos llamado MWA). Además se ha completado la Tabla 1 con una medida de dispersión como es la desviación típica del error obtenido (STD). No se ha tenido en cuenta el mes de diciembre ya que no suele ser representativo del comportamiento normal de la serie. En la Tabla 1 podemos observar que el modelo DHR produce unos errores de predicción (MWA) por debajo del 3% acompañado de una desviación típica muy baja. De la Tabla 1 se

corroborar como el DHR obtiene menor error sistemáticamente sobre el NAIVE además de menor dispersión.

**Tabla 1.** Errores de predicción en las demandas del 2002 del mercado español.

| Meses      | DHR     |      | NAIVE   |       |
|------------|---------|------|---------|-------|
|            | MWA (%) | STD  | MWA (%) | STD   |
| Enero      | 1.36    | 1.11 | 3.54    | 1.75  |
| Febrero    | 2.59    | 1.74 | 7.57    | 2.95  |
| Marzo      | 6.35    | 5.79 | 24.4    | 13.11 |
| Abril      | 2.59    | 2.02 | 18.3    | 10.12 |
| Mayo       | 2.27    | 1.52 | 15.5    | 8.15  |
| Junio      | 3.69    | 3.23 | 12.9    | 8.70  |
| Julio      | 2.33    | 1.82 | 9.38    | 7.50  |
| Agosto     | 2.22    | 1.61 | 15.1    | 9.96  |
| Septiembre | 3.04    | 2.58 | 14.1    | 8.92  |
| Octubre    | 2.20    | 1.94 | 10.9    | 3.62  |
| Noviembre  | 2.53    | 1.72 | 9.71    | 9.17  |

Hasta aquí hemos proporcionado un método que realiza predicciones de las demandas de energía de una manera sistemática, con una identificación objetiva y además el modelo es adaptativo, es decir, que no tenemos que actualizar el modelo cada cierto tiempo. No obstante, un paso más allá sería extender este modelo al caso de precios de la energía eléctrica. La serie temporal de precios es mucho más volátil que la formada por las demandas, esto hace que sea más difícil de predecir. No obstante es conocido que la serie temporal de precios comparte muchas características encontradas en el caso anterior, como es la no estacionariedad, la estacionalidad diaria y semanal, un rápido muestreo (también horaria), etc.

En la Tabla 2 hemos plasmado los errores de predicción calculados mediante el MWA y su desviación típica (STD) para los precios de energía correspondientes a la última semana de distintos meses procedentes del mercado español y del PJM interconnection (organización de transmisión regional que coordina la transacción de electricidad en total o parte de Delaware, Illinois, Maryland, New Jersey, Ohio, Pennsylvania, Virginia, West Virginia y el distrito de Columbia). Gracias a que hemos normalizado la medida de error (9) podemos comparar los errores cometidos en la serie de demandas con los errores de los precios. En dicha tabla podemos observar que los errores cometidos en la predicción de precios son del orden del 10% para el caso español y en torno al 15% para el PJM. De nuevo podemos comprobar que los errores son sistemáticamente más bajos en el modelo DHR que el modelo NAIVE, además de poseer menor dispersión.

#### 4. Paquete informático: Forecastool

Hasta ahora hemos presentado las características científicas del modelo DHR. Todo esto se engloba dentro del entorno de investigación. Sin embargo todavía falta conducir la investigación hacia la innovación. Para conseguir esto se ha implementado una herramienta informática llamada *forecastool* que pueda hacer posible que nuestro modelo se aplique en la realidad. Normalmente, los paquetes de software están escritos en lenguajes de programación estándar, como C++. Sin embargo esto puede conllevar un alto coste. Forecastool ha sido desarrollado dentro del entorno de MATLAB, usando m-files de MATLAB procedentes de la Toolbox de CAPTAIN, para implementar el modelo UC-DHR descrito en las secciones

anteriores. Nuestro objetivo es desarrollar una herramienta que pueda ser utilizada por una persona que no sea experta en series temporales. Por tanto, el programa va a ser fácil de usar ya que se ha utilizado un entorno amigable proporcionado por la *Graphical User Interface* (GUI) de MATLAB. Una herramienta parecida se desarrolló en Tych *et al.* (2002) para gestionar las llamadas telefónicas recibidas por una compañía, con lo que se dispone de una experiencia empresarial previa en la implantación de este tipo de software.

**Tabla 2.** Errores de predicción en los precios del 2002 del mercado español y del PJM

| Meses      | MERCADO ESPAÑOL |       |         |       | MERCADO PJM |       |         |       |
|------------|-----------------|-------|---------|-------|-------------|-------|---------|-------|
|            | DHR             |       | NAIVE   |       | DHR         |       | NAIVE   |       |
|            | MWA (%)         | STD   | MWA (%) | STD   | MWA (%)     | STD   | MWA (%) | STD   |
| Enero      | 16.01           | 14.36 | 31.06   | 24.17 | 12.15       | 12.20 | 51.42   | 54.37 |
| Febrero    | 7.64            | 5.29  | 8.36    | 4.927 | 12.97       | 16.35 | 49.38   | 52.48 |
| Marzo      | 9.53            | 8.78  | 27.86   | 18.77 | 16.87       | 14.49 | 50.33   | 54.89 |
| Abril      | 6.76            | 5.88  | 17.01   | 15.36 | 13.17       | 13.29 | 45.47   | 50.86 |
| Mayo       | 6.29            | 4.77  | 14.01   | 12.79 | 35.27       | 30.52 | 45.36   | 44.36 |
| Junio      | 10.64           | 9.95  | 26.98   | 18.50 | 23.84       | 22.47 | 47.23   | 47.19 |
| Julio      | 11.49           | 8.93  | 20.64   | 19.32 | 38.76       | 62.57 | 54.98   | 69.66 |
| Agosto     | 15.82           | 12.43 | 29.68   | 15.91 | 18.57       | 21.83 | 38.32   | 46.20 |
| Septiembre | 12.94           | 11.02 | 31.23   | 18.62 | 20.47       | 20.64 | 42.22   | 48.36 |
| Octubre    | 11.92           | 10.72 | 19.20   | 12.92 | 13.32       | 12.83 | 39.50   | 41.97 |
| Noviembre  | 14.10           | 12.99 | 43.29   | 19.49 | 16.02       | 18.43 | 48.69   | 53.32 |

Al inicio del artículo dijimos que además de presentar ventajas de carácter técnico, también proporcionamos ventajas de cara a la gestión de operaciones. Esto se debe a que nuestro modelo está enmarcado dentro del espacio de los estados. Lo que se traduce en conseguir extraer de la señal la componente de tendencia, la componente estacional y la componente irregular. Estas componentes pueden ser de gran ayuda a la hora de que los gerentes definan las estrategias que maximicen sus rentabilidades.

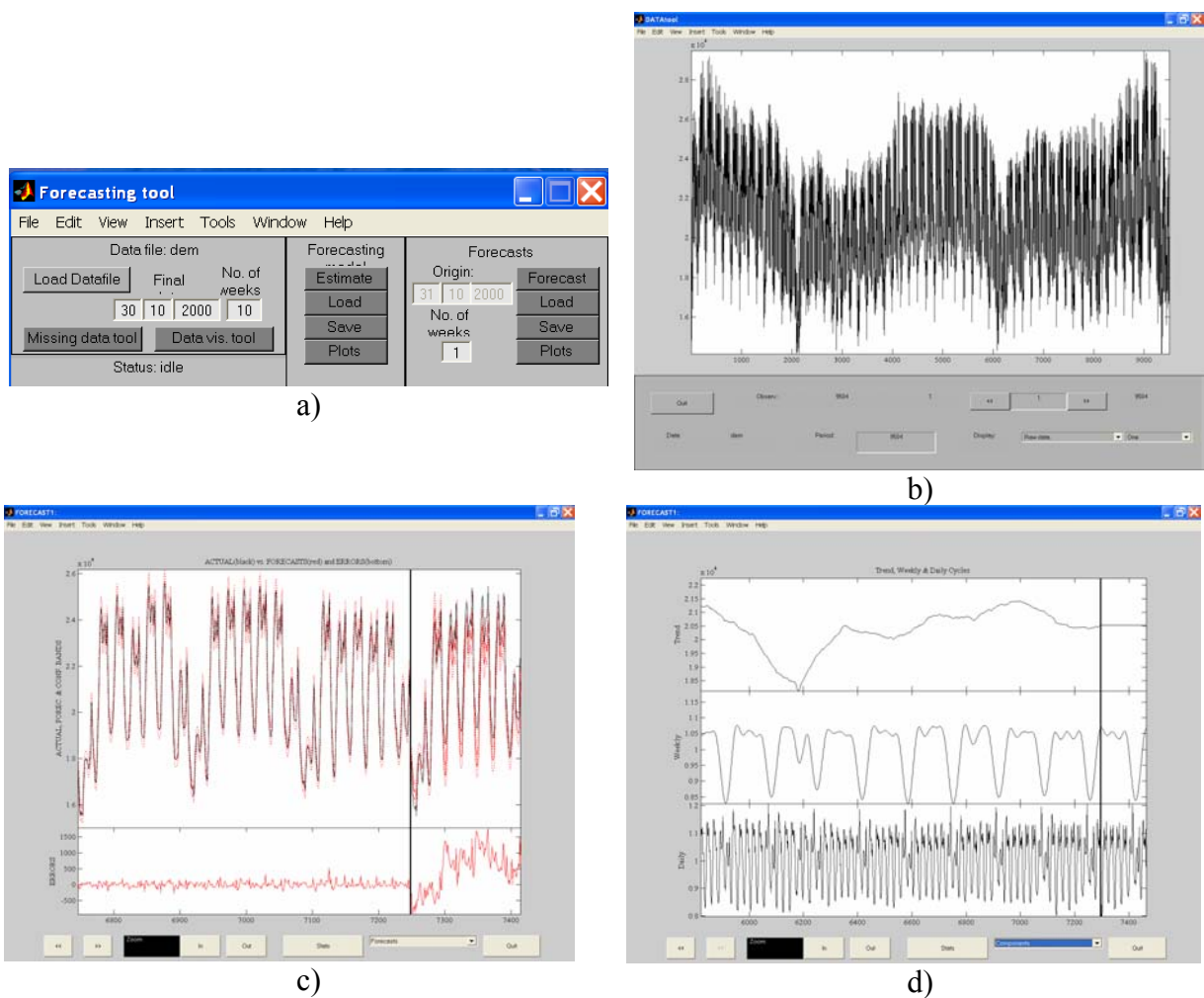
#### 4.1. Funcionamiento de forecastool

El paquete informático se divide en tres módulos mostrados en el panel principal de control, ver Figura 1 (a). El primer módulo realiza el análisis inicial de los datos. De manera que podemos cargar los datos, y mediante el botón de *Data Visualization Tool*, Figura 1 (a), abrimos otra ventana, Figura 1 (b), que nos permite visualizar los datos de diferentes maneras. Por ejemplo podemos segmentar los datos en función de la periodicidad introducida, superponerlos frente a la media y su desviación típica para la detección de atípicos, etc.

El segundo módulo realiza la estimación de los datos mediante los algoritmos KF/FIS. Las estimaciones las podemos guardar y dibujarlas.

El tercer módulo realiza la predicción. Solamente tenemos que definir el número de semanas que queremos predecir y darle al botón *Forecast* del panel principal. En la Figura 1 (c) podemos ver el resultado, donde se muestra las semanas utilizadas para la estimación y la predicción obtenida. En la parte inferior de la figura podemos observar los errores cometidos. El programa también nos permite obtener como salida gráfica la tendencia y la estacionalidad semanal y diaria tanto de la estimación como de la predicción. En la Figura 1 (d) podemos ver

la tendencia de la serie temporal de las demandas de energía eléctrica del mercado español. A su vez se muestra la estacionalidad diaria y la semanal.



**Figura 1.** a) Panel de control principal. b) *Data visualization tool*. c) Estimación y predicción. d) Componentes

## 5. Conclusiones

En el presente artículo se ha mostrado una herramienta de predicción a corto plazo, de precios y demandas de energía eléctrica, basada en un modelo de componentes no observables expresado en forma de espacio de los estados y actualizado mediante el filtro de kalman. Esta herramienta presenta varias ventajas que vamos a resumir:

- La identificación y estimación se realiza en el dominio de la frecuencia, dada la múltiple estacionalidad de las series estudiadas. Este dominio nos permite identificar de una manera objetiva las frecuencias a incluir, automatizando así la realización de dichas tareas.
- Debido a la utilización de un filtro adaptativo como es el filtro de kalman nos aseguramos que nuestro modelo se adapte a los cambios ocurridos en las series temporales.
- El marco de espacio de los estados empleado nos permite extraer información muy valiosa como puede ser la tendencia o la componente estacional de las series.



- d) Los errores de predicción calculados son muy competitivos. En la serie de demandas el orden del MWA suele ser menor del 3%, en el caso de los precios del mercado español son en torno al 10% y para los precios del PJM obtenemos errores del 10-20%.

La robustez de nuestra herramienta queda corroborada mediante su utilización en distintos mercados eléctricos. Todas estas características hacen nuestra herramienta muy atractiva de cara a facilitar la toma de decisiones de los diferentes agentes del mercado.

Por otra parte hemos mostrado la herramienta informática llamada *forecastool* encargada de llevar a cabo nuestro modelo de una manera fácil y sistemática. Con esto conseguimos que los gerentes encargados de realizar las predicciones puedan utilizar nuestro modelo sin tener que tener un conocimiento específico de él. Además proporcionamos la tendencia, sus componentes estacionales y el error cometido con el fin de que los gerentes puedan incluirlos en sus sistemas de información para realizar sus estrategias de oferta/compra.

Todas estas ventajas tanto técnicas como de gestión sitúan nuestra herramienta en una posición privilegiada para su aplicación en los mercados eléctricos.

## Apéndice

Los algoritmos del filtro de kalman y el suavizado de intervalo fijo aplicados a los casos de estudio en este artículo se pueden escribir en un formato general de espacio de los estados de la siguiente manera:

### 1. Filtro de kalman:

a) Ecuaciones de predicción:

$$\begin{aligned}\hat{\mathbf{x}}_{t|t-1} &= \mathbf{A}\hat{\mathbf{x}}_{t-1} \\ \hat{\mathbf{P}}_{t|t-1} &= \mathbf{A}\hat{\mathbf{P}}_{t-1}\mathbf{A}^T + \mathbf{Q}\end{aligned}$$

b) Ecuaciones de actualización:

$$\begin{aligned}F_t &= [1 + \mathbf{H}_t\hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T] \\ \hat{\mathbf{x}}_t &= \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T(z_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1})/F_t \\ \hat{\mathbf{P}}_t &= \hat{\mathbf{P}}_{t|t-1} - \hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T\mathbf{H}_t\hat{\mathbf{P}}_{t|t-1}/F_t\end{aligned}$$

### 2. Suavizado de intervalo fijo

$$\begin{aligned}\hat{\mathbf{x}}_{t|N} &= \hat{\mathbf{x}}_{t|t-1} + \hat{\mathbf{P}}_{t|t-1}\mathbf{s}_{t-1} \\ \hat{\mathbf{P}}_{t|N} &= \hat{\mathbf{P}}_{t|t-1} - \hat{\mathbf{P}}_{t|t-1}\mathbf{S}_{t-1}\hat{\mathbf{P}}_{t|t-1} \\ \mathbf{s}_{t-1} &= \mathbf{H}_t^T(z_t - \mathbf{H}_t\hat{\mathbf{x}}_{t|t-1})/F_t + \overline{\mathbf{\Phi}}_t^T\mathbf{s}_t \quad \text{with } \mathbf{s}_N = \mathbf{0} \\ \mathbf{S}_{t-1} &= \mathbf{H}_t^T\mathbf{H}_t/F_t + \overline{\mathbf{\Phi}}_t^T\mathbf{S}_t\overline{\mathbf{\Phi}}_t \quad \text{with } \mathbf{S}_N = \mathbf{0} \\ \overline{\mathbf{\Phi}}_t &= \mathbf{\Phi}_t - \mathbf{\Phi}_t\hat{\mathbf{P}}_{t|t-1}\mathbf{H}_t^T\mathbf{H}_t/F_t\end{aligned}$$

## Referencias

- Contreras, J; Espínola R.; Nogales, F.J.; Conejo A. (2003). ARIMA models to predict next-day electricity prices, *IEEE Trans. Power Syst.*, Vol. 18, pp. 495-500.
- Harvey, A. C. (1989) *Forecasting Structural Time Series Models and the Kalman Filter*, Cambridge: Cambridge University Press.
- Harvey, A. C.; Koopman, S. J. (1993). Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, vol. 88, pp. 1228-1242.
- Kumluca, A.; Erkmen, I. (2003), A hybrid learning for neural networks applied to short ter load forecasting, *Neurocomputing*, Vol. 51, pp.495-500.

Pedregal D. J.; Young, P. C. (2004). Modulated cycles, a new approach to modelling seasonal/cyclical behaviour in unobserved component models. Enviado a la *International Journal of Forecasting*. Una version previa está disponible como Tech. Note No. TR/145, *Centre for Research on Environmental Systems and Statistics (cres)*, Lancaster University (UK).

Pedregal D. J.; Young P. C. (2002) Statistical approaches to modelling and forecasting time series. In Clemets, M. and Hendry, D. (eds.). *Companion to Economic Forecasting*. Blackwell Publishers, pp. 69-104.

Trapero J. R. (2003), Modelización de mercados eléctricos mediante modelos de componentes no observables, Proyecto final de carrera. Área de organización de empresas, ETS Ingenieros Industriales, Universidad de Castilla-La Mancha.

Tych, W.; Pedregal D. J.; Young, P. C.; Davies. J. (2002) A software package for multi-rate unobserved component forecasting of telephone call demand, *International Journal of Forecasting*, vol. 18, pp. 673-695.

West, M.; Harrison, J. (1989) *Bayesian Forecasting and Dynamic Models*, New Cork: Springer-Verlag.

Young, P.C. (1994) Time-variable parameter and trend estimation in Nonstationary economic time series. *Journal of Forecasting*, vol. 13, pp. 179-210.

Young, P.C.; Pedregal D.J.; Tych, W. (1999), Dynamic Harmonic Regression, *Journal of Forecasting*, vol. 18, pp. 369-394.

Young, P.C.; Taylor, J. C.; Pedregal, D. J.; Tych, W.; McKenna, P. G. (2003); *System Identification, Time Series Analysis and Forecasting. The Captain Toolbox. Centre for Research on Environmental Systems and Statistics*, Lancaster University. Disponible: <http://www.es.lacs.ac.uk/cres/captain>